DMAP: integrated mobility and service management in mobile IPv6 systems

Ing-Ray Chen · Weiping He · Baoshan Gu

Abstract Mobile IPv6 (MIPv6) is a work in progress IETF standard for enabling mobility in IPv6 networks and is expected to have wide deployment. We investigate an integrated mobility and service management scheme based on MIPv6 with the goal to minimize the overall network signaling cost in MIPv6 systems for serving mobility and service management related operations. Our design extends IETF work-in-progress Hierarchical Mobile IPv6 (HMIPv6) with the notion of dynamic mobility anchor points (DMAPs) for each mobile node (MN) instead of static ones for all MNs. These DMAPs are access routers chosen by individual MNs to act as a regional router to reduce the signaling overhead for intra-regional movements. The DMAP domain size, i.e., the number of subnets covered by a DMAP, is based on the MN's mobility and service characteristics. Under our DMAP protocol, a MN interacts with its home agent and application servers as in the MIPv6 protocol, but optimally determines when and where to launch a DMAP to minimize the network cost in serving the user's mobility and service management operations. We demonstrate that our DMAP protocol for integrated mobility and service management yields significantly improved performance over basic MIPv6 and HMIPv6.

I-R. Chen (✉) · W. He · B. Gu
Department of Computer Science, Virginia Tech, 7054
Haycock Road, Northern Virginia Center, Falls
Church, VA 22043, USA
e-mail: irchen@vt.edu

W. He
e-mail: weiping@vt.edu

B. Gu
e-mail: bgu@vt.edu

1 Introduction

IPv6 networks have been deployed rapidly all over the world. Mobile IPv6 (MIPv6) [5] is a network protocol for enabling mobility in IPv6 networks. It allows mobile nodes (MNs) to move within IP-based networks while maintaining on-going connections. MIPv6 has been flagged as the mobility management protocol for future all-IP mobile systems and is expected to have wide deployment. With the anticipated increase in inexpensive, computationally powerful mobile devices running mobile applications to access multimedia and data services over broadband wireless connections based on IPv6, it is of fundamental importance to research effective and scalable network management

schemes based on MIPv6 to reduce network costs in IPv6 networks.

In this paper, we investigate a scalable and efficient integrated mobility and service management scheme, called DMAP, based on MIPv6 with the goal to minimize the network cost in MIPv6 systems for serving mobility and service management related operations. Our design extends IETF work-in-progress Hierarchical Mobile IPv6 (HMIPv6) [10] with the notion of "dynamic" mobility anchor points (DMAPs) for each individual MN instead of static ones for all MNs. These DMAPs are simply access routers (ARs) chosen by individual MNs to act as a regional router to reduce the signaling overhead for intra-regional movements. The DMAP domain size, or the number of subnets in a region covered by a DMAP, is based on the mobility and service characteristics of each MN.

We aim at the design of DMAP to integrate seamlessly with MIPv6. The goal is to identify the optimal DMAP domain size dynamically to minimize the network communication overhead due to mobility and service operations induced by each MN. When a MN enters a subnet, the MN determines if the AR would be selected as the DMAP. If yes, the MN obtains a regional care of address (RCoA) from the DMAP and informs its home agent (HA) and application servers of the RCoA address change. Essentially with DMAP, a MN interacts with the HA and application servers following the standard MIPv6 protocol but dynamically determines when and where to launch a DMAP so as to minimize the network cost.

Our proposed DMAP impacts the MIPv6 Internet standard and benefits IPv6 networks by effectively integrating mobility and service management to reduce networking costs. DMAP determines the optimal MAP domain size to minimize network costs for both mobility and service induced management operations. We demonstrate the network cost saving to be significantly better than that provided by basic MIPv6 and HMIPv6 [10] which deal with mobility management only. Furthermore, the network cost reduction benefit as a result of applying the proposed scheme would be cumulative and proportional to the number of MNs. In this paper, we will use the term "service area" interchangeably with "MAP domain" since in our design we deal with service management in addition to mobility

management. A location handoff refers to the MN moving across a subnet boundary, while a service handoff refers to the MN moving across a DMAP domain boundary.

The rest of the paper is organized as follows. Section 2 describes related work. Section 3 describes our proposed DMAP scheme for integrated mobility and service management in MIPv6 environments. Section 4 develops a computational procedure to determine the optimal service area for service handoffs in order to minimize the network communication cost induced by mobility and service management operations. In Section 5, we compare DMAP with basic MIPv6 and HMIPv6. Finally, Section 6 summaries the paper and outlines some future research areas.

2 Related work

For next-generation mobile IPv6 networks, MNs are expected to be very active with significant mobility. The mobility rate with which subnets are crossed by MNs can be high, causing a high signaling overhead for the MN to inform HA and CNs of the CoA address change. There have been approaches [2,7,10,12,13] proposed to mitigate this high volume of network signaling cost, including, most noticeably, IETF work-in-progress MIP Regional Registration (MIP-RR) [4], Hierarchical MIPv6 [10] and IDMP [7]. MIP-RR uses a Gateway Foreign Agent (GFA) to provide a regional CoA, which acts as a proxy for regional movement management to keep track of the MN's current CoA as long as the MN moves within a region, thereby reducing the network signaling cost when the MN moves within a region. When the MN moves to a new region, the regional CoA change is informed to the HA. The design is for mobility management only without considering service management-induced network cost.

Hierarchical MIPv6 (HMIPv6) [10] is designed to reduce the network signaling cost for mobility management based on the observation that statistically local mobility accounts for more than 60% of movements made by a MN. In addition to a CoA, a regional CoA (RCoA) is also allocated to a MN whenever the MN enters a new MAP domain. The HA and CNs ideally only know the

MN's RCoA, so whenever the MN moves across a MAP domain and triggers a RCoA address change, the new RCoA address needs to be propagated to the HA and CNs. Whenever, a MN moves from one subnet to another but is still within a region covered by a MAP domain, the CoA change is only propagated to the MAP instead of to the HA and CNs, thus saving the signaling cost for mobility management. The number of subnets covered by a MAP domain is static in HMIPv6. This concept can be applied at multiple levels in a hierarchical manner [9,10] and can be combined with the use of Fast Handovers [2] that use forwarding pointers between the current and next subnets in a hybrid manner. MAPs in HMIPv6 are statically configured and shared by all MNs in the system. ARs are responsible for announcing their MAP's identity by means of router advertisement packets so that roaming MNs would know if they have crossed a MAP domain and need to perform a RCoA update to the HA and CNs. Our proposed DMAP extends HMIPv6 by taking both mobility and service management into consideration, such that each MN discriminatively selects ARs as MAPs and dynamically changes the size of a MAP domain (corresponding to a "service area") based on its mobility and service characteristics at runtime so as to minimize the total cost due to the signaling cost associated with mobility management and the network communication cost associated with service management.

Another IETF work-in-progress draft IDMP [7] introduces the concept of domain mobility with a domain corresponding to a region in MIP-RR and HMIPv6, and a domain agent corresponding to a MAP in HMIPv6 to keep track of CoA of a MN as the MN roams within a domain. IDMP can be combined with fast handoff mechanisms utilizing multicasting [6] to reduce handoff latency and paging mechanisms to reduce the network signaling cost for intra-domain movements.

HMIPv6 and IDMP can effectively reduce network signaling cost for mobility management compared with basic MIPv6 and are scalable to a large number of MNs. However, there is no mechanism provided to determine the size of a MAP domain in HMIPv6 (or a mobility domain in IDMP) for all MNs that would minimize the network cost, since they deal with mobility management only without

considering service characteristics of individual MNs. Our proposed DMAP scheme extends HMIPv6 by considering both mobility and service management. It aims to determine the best MAP domain size on a per MN basis in order to reduce the network signaling and communication cost resulting from both mobility and service management for future IP-based Internet applications.

3 DMAP for integrated mobility and service management

We propose DMAP that extends HMIPv6 for integrated mobility and service management to reduce the network signaling and communication overhead for servicing mobility and service induced operations. DMAP takes a step further beyond basic MIPv6 and HMIPv6 which we consider as special cases designed for mobility management only. The essence of DMAP is the notion of integrated mobility and service management. This is achieved by determining an optimal service area size corresponding to the size of a MAP domain in HMIPv6. This service area size is expressed in terms of the number of subnets. It is dynamically determined for each MN, taking into consideration of both mobility and service characteristics of the MN. We develop a computational procedure by which the optimal service area size can be determined. The objective is to minimize the total network signaling and communication overhead in servicing the MN's mobility and service management operations.

We implement DMAP by means of a "DMAP table lookup" design described as follows, leveraging binding request messages defined in MIPv6 and HMIPv6. This design has the advantage of simplicity, scalability and efficiency and is HMIPv6-compliant. When a MN crosses a service area, it makes the AR of the subnet just crossed as the DMAP as in HMIPv6. The MN also determines the size of the new service area (or MAP domain). Concurrently, it acquires a RCoA as well as a CoA from the current subnet and registers the address pair (RCoA, CoA) to the current DMAP (the AR of the current subnet) in a binding request message. Note that the RCoA could be the same as the CoA upon the MN's entry into a new DMAP

domain. The MN also informs the HA and CNs of the new RCoA address change in another binding message so that the HA and CNs would know the MN by its new RCoA address. When the HA and CNs subsequently send packets to the MN, they would use the RCoA as the MN's address.

A packet destined for RCoA will first be intercepted by the DMAP. By inspecting the address pair (RCoA, CoA) stored in the internal table, the DMAP knows that the MN's address is actually the CoA and will forward the packet to the MN through tunneling. If the RCoA and CoA are in the same subnet, the DMAP can directly forward the packet to the MN without using tunneling. When the MN subsequently crosses a subnet but is still located within the service area, it would inform the MAP of the CoA address change without informing the HA and CNs to reduce the network signaling cost. This "DMAP table lookup" design maps RCoA to CoA by having the current DMAP maintain an internal table, so the DMAP can intercept a packet destined for RCoA and forward it to the MN's CoA. It is efficient since the RCoA-CoA routing function can be performed efficiently by DMAPs (which are routers) through simple table lookup operations. It is scalable because the design is scalable to a large number of MNs by having all ARs in MIPv6 networks DMAP-enabled and randomly spreading the routing and table lookup functions to all ARs in the network. In terms of security and fault tolerance, it can also leverage
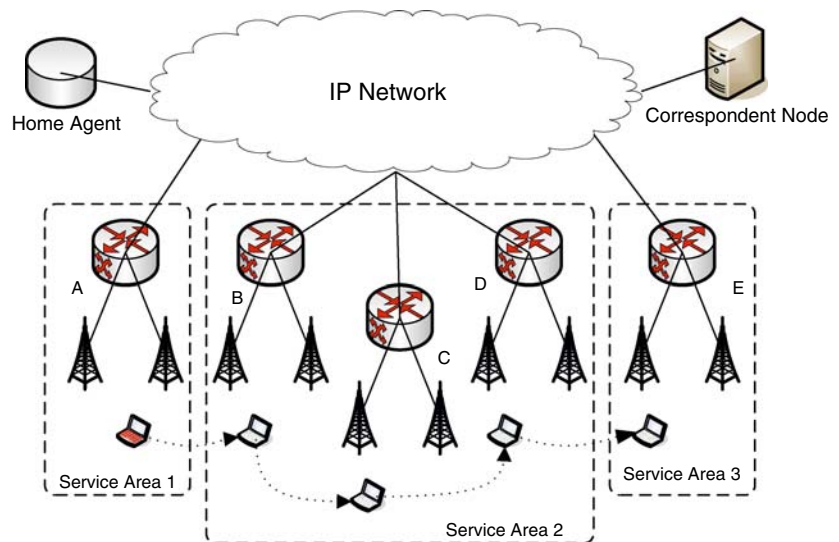
existing solutions in HMIPv6 because this design is HMIPv6-compliant except that a MN dynamically selects ARs to be MAPs.

The concept of per-user service DMAP derives from our earlier work on service management in wireless personal communication networks [3]. The idea of determining a dynamic "service area" to minimize the network signaling cost is similar in concept to determining a gateway foreign agent (GFA) coverage area in MIP Regional Registration [12]. However, MIP Regional Registration only deals with mobility management while our scheme deals with both mobility and service management.

A MN's service area can be modeled as consisting of K IP subnets. We develop a computational procedure to determine the optimal service area size in terms of K. The optimal K value would be determined at runtime dynamically to minimize the network cost. For the special case in which K is constant for all MNs, our scheme degenerates to HMIPv6 with a two-level MAP-AR structure.

Figure 1 illustrates DMAP in MIPv6 environments. When moving from one subnet and another within service area 1, the MN only informs its CoA change to the DMAP without informing the HA or CNs. When the MN exits service area 1, the AR of subnet B becomes the DMAP. The MN obtains a CoA and a RCoA from subnet B and an entry (RCoA, CoA) is recorded in the routing table of the AR of subnet B. Subnet B's AR now acts as

Fig. 1 DMAP: integrated mobility and service management in MIPv6

the DMAP of the MN. Both the HA and CNs are informed of the RCoA address change by the MN.

In our DMAP scheme, the MN appoints a new DMAP only when it crosses a service area whose size is determined based on the mobility and service characteristics of the MN in the new service area. One should note that the service area size of the DMAP is not necessarily uniform. In the above scenario although subnet B appears to be at the edge of service area 2, it is actually at the center of the new service area since our service area is defined by the number of subnets (or the number of moves) starting from the first subnet at which the MN enters into a new service area. Within service area 2 (the new service area that the MN just moves into), if the MN moves from subnet B to subnet D through C (with 2 location handoffs), then the DMAP will be informed of the CoA change by the MN but will remain at the same location (subnet B).

The optimal service area size is dictated by the MN's mobility and service characteristics. A large service area size means that the DMAP will not change often. The consequence of not changing the DMAP often is that the service delivery cost would be high because of the triangular routing path CN-DMAP-MN for data communication between the CN and MN. On the other hand, a small service area size means that the DMAP will be changed often so it will stay close to the MN. The consequence is that the communication cost for service data delivery would be low because of the short CN-DMAP-MN route. However, a DMAP change involves the cost of informing the HA and CNs of the RCoA address change. Therefore, there is a trade-off between these two cost factors and an optimal service area exists.

The service and mobility characteristics of a MN are summarized by two parameters. The first parameter is the residence time that the MN stays in a subnet. This parameter can be collected by each MN based on statistical analysis [1]. We expect that future MNs are reasonably powerful for collecting data and doing statistical analysis. The residence time in general would be characterized by a general distribution. Loosely, we use the MN's mobility rate (σ) to represent this parameter. The second parameter is the service traffic between the MN and server applications. The MN can also collect data statistically to parameterize this. Loosely, we use the data packet rate (λ) between the MN and CNs to represent this parameter. Both of these parameters are to be determined by the MN. For efficiency, the MN could build a table to lookup its mobility and service rates as a function of its location, time of the day, and day of the week, based on statistical analysis. The ratio of λ/μ is called the service to mobility ratio (SMR) of the MN.

When a MN moves across a service area boundary, the DMAP changes, thus incurring a "service handoff." This service handoff cost includes the signaling cost to inform the HA and CNs of the new RCoA of the MN. In IPv6-based systems, the communication overhead between two communicating processes is measured by the number of hops. Since the number of subnets separating two communicating processes would not properly measure the hop-count distance, we let $F(K)$ denote a function that returns the number of hops as a function of the number of subnets K. This function $F(K)$ can be periodically and dynamically determined by a MN which collects statistical data as it roams across subnets. A MN would determine this function dynamically. The fluid flow model [13] assumes that the average number of hops between two communicating processes separated by K subnets is equal to $\sqrt{K}$.

Table 1 lists a set of identified system parameters that characterize the mobility and service characteristics of a MN in a MIPv6 system.

Our DMAP scheme is per-user based. Data packets from a server application are sent to a MN's DMAP first (which is just an AR selected to be the current DMAP) and then forwarded to the MN. The DMAP will receive packets addressed to the RCoA from the HA and CNs. Packets will be tunneled from the DMAP to the MN.

3.1 Intra-regional move

When a MN performs a location handoff within a service area, it acquires a CoA from the subnet and informs the DMAP of the CoA address change. The DMAP is not changed. Also, the HA and CN are not informed of the CoA address change of the MN since they only know the MN by the RCoA address which is not changed in this case.

Table 1 Parameters for integrated mobility and service management

Symbol	Meaning
λ	Data packet rate between the MN and CNs
σ	Mobility rate at which the MN moves across subnet boundaries
SMR	Service to mobility ratio (λ/σ)
N	Number of server engaged by the MN
$F(K)$	A general function relating the number of subnets K to the number of hops
K	Number of subnets in one service area
τ	1-hop communication delay per packet in wired networks
α	Average distance between HA and MAP
β	Average distance between CN and MAP
γ	Cost ratio between wireless vs. wired network

3.2 Inter-regional move

When a MN moves across a service area thus incurring a service handoff, the MN acquires a CoA and a RCoA from the AR that now becomes the DMAP and an entry (RCoA, CoA) is recorded in the lookup table of the AR. The MN then informs the HA and CNs of its new RCoA to complete the service handoff. The implementation of the proposed DMAP scheme based on the "DMAP table lookup" design is totally transparent to the HA and CNs. The HA and CNs are informed of the RCoA as part of the service handoff process. Packets from the HA or CNs will use the RCoA as the destination address, which will be intercepted by the AR that serves as the MN's DMAP who will then forward them to the MN. The packet routing and forwarding will be done entirely in the network layer. In effect the system behaves as if a two-level HMIPv6 structure has been used to do both mobility and service management, except that the service area is to be determined by the MN. Instead of having the MN discover the presence of a MAP through announcement packets and initiating the MAP migration process, the MN will dynamically determine which AR will act as a DMAP to minimize the network cost.

4 Model

We devise a computational procedure to determine the optimal service area size utilizing stochastic Petri net (SPN) techniques. The intent to find the optimal service area based on the MN's mobility and service behaviors. The designer would utilize the computational procedure developed here to build a table at static time listing the optimal service area as a function of these parameters each covering a reasonable value range. Such a table is then loaded into the MN. The actual values of these parameters are dynamically collected by the MN at runtime. Based on the values of these parameters at the time a service area is crossed, the MN performs a simple table lookup to determine the optimal service area. The computational procedure requires that (1) every AR must be capable of acting as a mobile anchor point (MAP); and (2) each MN must be powerful enough to collect data dynamically and perform simple statistical analysis to parameterize its mobility and service characteristics.

The metric that we aim to minimize is the "communication cost" incurred per time unit due to mobility and service operations. The communication cost includes the signaling overhead for mobility management for informing the DMAP of the CoA changes, and informing the HA and CNs of the RCoA changes, as well as the communication overhead for service management for delivering data packets between the MN and CNs. Our SPN model is shown in Fig. 2. Table 2 gives the meaning of places and transitions defined in the model. We choose SPN because of its ability to deal with general time distributions for events, its concise representation of the underlying state machine to deal with a large number of states, and its expressiveness to reason about a MN's behavior as it migrates from one state to another in response to events occurring in the system. Once the parameters of the SPN model are given proper values, numerical
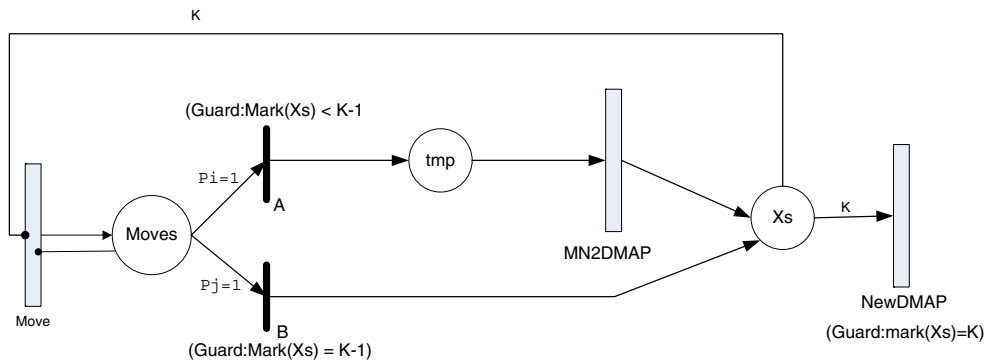
Fig. 2 Stochastic Petri net model

Table 2 Meaning of places and transitions in the SPN model

Symbol	Meaning
Move	A timed transition for the MN to move across subnet areas
Moves	Mark(Moves) = 1 meaning that the MN just moves across a subnet
MN2DMAP	A timed transition for the MN to inform the DMAP of the CoA change
Xs	Mark(Xs) holds the number of subnets crossed in a service area
NewDMAP	A timed transition to inform the HA and CNs of the RCoA change
K	Number of subnets under a service area
Guard:Mark(Xs)< $K-1$	A guard for transition A that is enabled if a move will not cross a service area
Guard:Mark(Xs)= $K-1$	A guard for transition B that is enabled if a move will cross a service area
Guard:Mark(Xs)= K	A guard for transition NewDMAP that is enabled if the MN just moves across a service area
A	An immediate transition when Mark(Xs)< $K-1$
B	An immediate transition when Mark(Xs)= $K-1$
tmp	A temporary place that holds tokens from transition A

analysis methods for solving SPN models based on SOR or Gauss Seidel [11] are readily available to compute the optimal service area size.

A token in the SPN model represents a subnet crossing event by the MN. The function Mark(P) returns the number of tokens in place P. The number of tokens accumulated in place Xs, that is, Mark(Xs), represents the number of subnets already crossed by the MN since it enters a new service area. The SPN model describes the behavior of a MN operating under the DMAP scheme:

1. When a MN moves across a subnet area, thus incurring a location handoff, a token is put in place Moves. The mobility rate at which location handoffs occur is σ which is the transition rate assigned to Move.
2. If the current move is an intra-regional move such that the guard for transition A will return *true*, then the MN will only inform the DMAP of the CoA change. This is modeled by firing immediate transition A, allowing the token in place Moves to move to place tmp. Subsequently, once the MN obtains a CoA from the subnet it just enters, it will communicate with the DMAP of the new CoA change. This is modeled by enabling and firing transition MN2DMAP. After MN2DMAP is fired, a token in place tmp flows to place Xs, representing that a location handoff has been completed and the DMAP has been informed of the CoA change of the MN.
3. If the current move results in the total number of moves being K such that the guard for transition B will return *true*, then the move will make the MN cross a service area. This is modeled by enabling and thus firing immediate transition B, allowing the token in place Moves to move to place Xs in preparation for a service handoff event. Note that in an SPN, firing an immediate transition does not take any time.

4. If the number of moves, including the current one, in place Xs has accumulated to K, a threshold determined by the DMAP representing the size of a service area, then it means that the MN has just moved into a new service area and a service handoff ensues. This is modeled by assigning an enabling function that will enable transition NewDMAP when K tokens have been accumulated in place Xs. After transition NewDMAP is fired, all K tokens are consumed and place Xs contains no token, representing that the AR of the subnet that the MN just enters has been appointed as the DMAP by the MN in the new service area.

Below we show an example of parameterizing transition rates of MN2DMAP and NewDMAP based on the set of base parameters defined in Table 1. The firing time of transition MN2DMAP stands for the communication time of the MN informing the DMAP of the new CoA through the wireless network. This time depends on the number of hops separating the MN and its DMAP. Thus, the transition rate of transition MN2DMAP is calculated as:

$$\frac{1}{\gamma\tau + F(\mathrm{Mark}(\mathrm{Xs}) + 1) \times \tau},$$

where τ stands for the one-hop communication delay per packet in the wired network and γ is a proportionality constant representing the ratio of the communication delay in the wireless network to the communication delay in the wired network. $F(\mathrm{Mark}(\mathrm{Xs})+1)$ returns the number of hops between the current subnet and the DMAP separated by $\mathrm{Mark}(\mathrm{Xs})+1$ subnets. The argument of the $F(x)$ function is added by 1 to satisfy the initial condition that $\mathrm{Mark}(\mathrm{Xs}) = 0$ in which the DMAP has just moved into a new service area, so at the first subnet crossing event, the distance between the DMAP and the subnet is one subnet apart. Note that this transition rate is state-dependent because the number of tokens in place Xs changes dynamically over time. When transition NewDMAP fires, the AR of the subnet that the MN moves into will be selected as the DMAP. The communication cost includes that for the MN to inform the HA and CNs of the new RCoA address change, i.e., $(\alpha + N\beta)\tau$, where α is the average distance in hops between the MN and the HA, β is the average distance

in hops between the MN and a CN, and N is the number of CNs that the MN concurrently engages. Thus, the transition rate of transition NewDMAP is calculated as:

$$\frac{1}{(\alpha + N\beta)\tau}.$$

The stochastic model underlying the SPN model is a continuous-time semi-Markov chain with the state representation of (a, b) where a is the number of tokens in place Moves, b is the number of tokens in place Xs. Let P_i be the steady state probability that the system is found to contain i tokens in place Xs such that $\mathrm{Mark}(\mathrm{Xs})=i$. Let $C_{i,\mathrm{service}}$ be the communication overhead for the network to service a data packet when the MN is in the ith subnet in the service area.

Let C_{service} be the average communication overhead to service a data packet weighted by the respective P_i probabilities. The communication overhead includes a communication delay between the DMAP and a CN in the fixed network, a delay from DMAP to the AR of the MN's current subnet in the fixed network, and a delay in the wireless link from the AR to the MN. Thus, C_{service} is calculated as follows:

$$C_{\mathrm{service}} = \sum_{i=0}^{K}(P_i \times C_{i,\mathrm{service}})$$

$$= \gamma\tau + \beta\tau + \sum_{i=0}^{K}(P_i \times F(i)\tau). \tag{1}$$

Let $C_{i,\mathrm{location}}$ be the network signaling overhead to service a location handoff operation given that the MN is in the ith subnet in the service area. If $i < K$, only a minimum signaling cost will incurred for the MN to inform the DMAP of the CoA address change. On the other hand, if $i = K$, then the location handoff also triggers a service handoff. A service handoff will incur a higher communication signaling cost to inform the HA and N CNs (or application servers) of the RCoA address change. Let C_{location} be the average communication cost to service a move operation by the MN weighted by the respective P_i probabilities. Then, C_{location} is calculated as follows:

$$C_{\text{location}} = \sum_{i=0}^{K}(P_i \times C_{i,\text{location}})$$
$$= P_K(\gamma\tau + \alpha\tau + N\beta\tau)$$
$$+ \sum_{i=0}^{K-1}\{P_i(\gamma\tau + F(K)\tau)\}. \qquad (2)$$

Summarizing above, the total communication cost *per time unit* for the Mobile IP network operating under our DMAP scheme to service operations associated with mobility and service management of the MN is calculated as:

$$C_{\text{DMAP}} = C_{\text{service}} \times \lambda + C_{\text{location}} \times \sigma, \qquad (3)$$

where λ is the data packet rate between the MN and CNs, and σ is the MN's mobility rate. The steady-state probability P_i, $1 \leq i \leq K$, needed in Eqs. 1 and 2 can be solved easily utilizing numerical method solution techniques such as SOR or Gauss Seidel [11].

5 Numerical results

Here we apply Eqs. 1–3 to calculate C_{DMAP} as a function of K and determine the optimal K representing the optimal "service area" size that will minimize the network cost, when given a set of parameter values characterizing the MN's mobility and service behaviors. Below we present results to show that there exists an optimal service area for systems operating under DMAP for network cost minimization, and demonstrate the benefit of DMAP over basic MIPv6 and HMIPv6.

For basic MIPv6, there is no DMAP. Thus, the communication cost $C_{\text{service}}^{\text{MIPv6}}$ for servicing a packet delivery in basic MIPv6 includes a communication delay from the CN to the AR of the current subnet, and a delay in the wireless link from the AR to the MN as follows:

$$C_{\text{service}}^{\text{MIPv6}} = \gamma\tau + \alpha\tau. \qquad (4)$$

Under basic MIPv6, when a MN crosses a subnet boundary, thus incurring a location handoff, the MN informs the HA and CNs of its CoA change. The cost $C_{\text{location}}^{\text{MIPv6}}$ for servicing a location handoff under basic MIPv6 consists of a delay in the wireless link from the MN to the AR of the subnet that it just enters into, a delay from that AR to the CNs, and a delay from that AR to the HA as follows:

$$C_{\text{location}}^{\text{MIPv6}} = \gamma\tau + \alpha\tau + N\beta\tau. \qquad (5)$$

The total cost per time unit for servicing data delivery and mobility management operations under MIPv6 is given by:

$$C_{\text{MIPv6}} = C_{\text{service}}^{\text{MIPv6}} \times \lambda + C_{\text{location}}^{\text{MIPv6}} \times \sigma. \qquad (6)$$

For HMIPv6, the placement of MAPs is predetermined. We compare DMAP with an implementation of HMIPv6 in which each MAP covers a fixed number of subnets, say, $K_H = 4$. Thus, when a MN crosses a subnet within a MAP domain of K_H subnets, it only informs the MAP of its CoA. On the other hand, when the MN crosses a MAP domain, it changes the MAP, obtains a new RCoA and informs the HA and CNs of the new RCoA.

Table 3 compares the communication cost incurred per time unit by DMAP versus that by basic MIPv6 and HMIPv6 head-to-head as a function of SMR, from the perspective of K_{opt} that minimizes the overall communication cost. The example is based on $F(K) = \sqrt{K}$ [13], $\alpha = \beta = 30$, and $\gamma = 10$. The cost metric is normalized with respect to $\tau = 1$. First we observe that K_{opt} exists under DMAP. Furthermore, as SMR increases,

Table 3 Comparing DMAP with basic MIPv6 and HMIPv6 head-to-head from the perspective of K_{opt}

SMR	C_{MIPv6}	$C_{\text{HMIPv6}}(K_H = 4)$	$C_{\text{DMAP}(K_{\text{opt}})}$	K_{opt}
0.1250	1.8750	0.7897	0.5522	34
0.2500	2.0000	0.9187	0.6892	32
0.5000	2.2500	1.1766	0.9619	28
1.0000	2.7500	1.6925	1.5034	22
2.0000	3.7500	2.7242	2.5758	16
4.0000	5.7500	4.7876	4.6960	11
8.0000	9.7500	8.9144	8.8859	7
16.0000	**17.7500**	**17.1681**	**17.1681**	**4**
32.0000	33.7500	33.6754	33.5475	2
64.0000	65.7500	66.6901	65.7500	1

K_{opt} decreases because when SMR is large, the service rate is high compared with the mobility rate, so the DMAP likes to stay as close to the MN as possible to minimize the communication overhead through the CN-DMAP-MN route. This table clearly exhibits the behavior of DMAP. We see that DMAP dominates basic MIPv6 when SMR is low. As SMR increases exceeding a threshold (e.g., 64), K_{opt} approaches 1 under which DMAP degenerates to basic MIPv6. The reason is that when SMR is sufficiently high, the MN's packet arrival rate is much higher than the mobility rate, so the data delivery cost dominates the mobility management cost. Therefore, the DMAP will stay close to the MN to lower the data delivery cost, thus making $K_{opt} = 1$ in our DMAP scheme in order to reduce the CN-DMAP-MN (or CN-MAP-MN) triangular routing cost for packet delivery.

Comparing DMAP with HMIPv6 from the perspective of K_{opt}, we observe from Table 3 that DMAP degenerates to HMIPv6 (with a fixed K_H) when SMR is moderate. When SMR is either high or low, DMAP performs substantially better than HMIPv6. The trend would be true with other choices of K_H for HMIPv6 as well (say $K_H = 8$). When SMR is low, K_{opt} is high for DMAP compared with a fixed K_H for HMIPv6, in which case the cost saving is due to mobility cost reduction. On the other hand, when SMR is high, K_{opt} is low for DMAP compared with a fixed K_H for HMIPv6, in which case the cost saving is due to service cost reduction. Overall, there is a wide range of SMR values under which DMAP dominates HMIPv6 because of the choice of K_{opt} for DMAP vs. a fixed K_H for HMIPv6. Here we should note that a small difference of 0.1 is considered significant because the metric is expressed in terms of the network cost per time unit (normalized to the per-hop cost $\tau = 1$) so the cumulative effect would be significant over the lifetime of a single MN. It is also worth noting that the cumulative effect of cost saving for a large number of MN would be even more significant.

Figure 3 summarizes the cost difference between basic MIPv6 and DMAP versus HMIPv6 and DMAP as a function of SMR. The Y coordinate is the cost difference incurred per time unit. There are two curves shown in Fig. 3. The first curve shows the cost difference between basic MIPv6

and DMAP ($C_{MIPv6} - C_{DMAP}$), and the second curve shows the cost difference between HMIPv6 and DMAP ($C_{HMIPv6} - C_{DMAP}$), as a function of SMR. We first observe that the cost difference between basic MIPv6 and DMAP (the first curve) decreases as SMR increases. The reason is that DMAP degenerates to MIPv6 when SMR becomes sufficiently large. We conclude that DMAP performs significantly better than basic MIPv6 especially when SMR is low. Next we observe that the cost difference between HMIPv6 and DMAP (the second curve) initially decreases as SMR increases until K_{opt} coincides with K_H at which point DMAP degenerates to HMIPv6, and then the cost difference increases sharply as SMR continues to increase. We conclude that DMAP performs significantly better than HMIPv6 when SMR is either low or high.

Figure 4 tests the sensitivity of the results with respect to α and β representing the hop distances between the MN and the HA and CNs. We see that as the hop distance between the HA (and CNs) increases, the cost difference between HMIPv6 and DMAP ($C_{HMIPv6} - C_{DMAP}$) becomes more pronounced, especially when SMR is small. The reason is that when SMR is small, the mobility management cost dominates the data delivery cost, so K_{opt} tends to be large to reduce the mobility management cost. In this case DMAP dictates a high K_{opt} value to be used to reduce the mobility management cost as opposed to a fixed $K_H = 4$ used by HMIPv6, thereby resulting in a more substantial cost difference between HMIPv6 and DMAP as SMR decreases. Recall that the cost is normalized with respect the per-hop communication cost $\tau = 1$. By observing the four curves in Fig. 4, nevertheless, we see that the trend remains the same. That is, the cost difference ($C_{HMIPv6} - C_{DMAP}$) decreases with the increase of SMR and then increases sharply after a threshold point at which DMAP degenerates to HMIPv6. Figure 4 shows that under a wide range of α- and β- values, DMAP always incurs less network overheads than HMIPv6, the effect of which is pronounced when SMR is relatively low or high.

All the above results are based on the assumption that the average number of hops between the DMAP and MN separated by k subnets is given by $F(k) = \sqrt{k}$, which are based on the fluid flow model

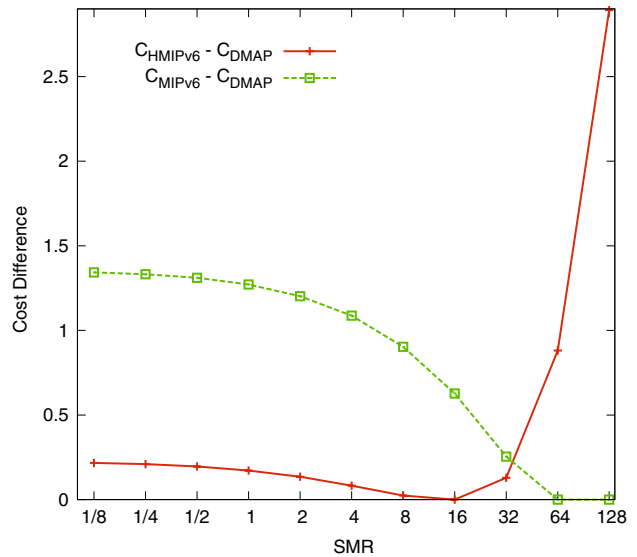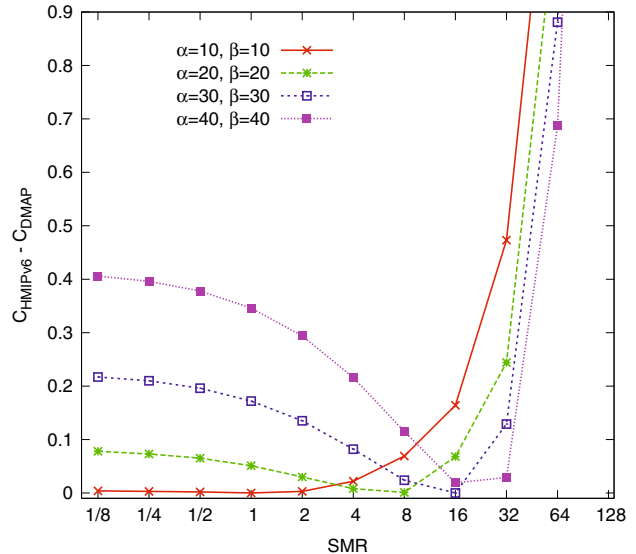**Fig. 3** Cost difference between basic MIPv6, HMIPv6, and DMAP

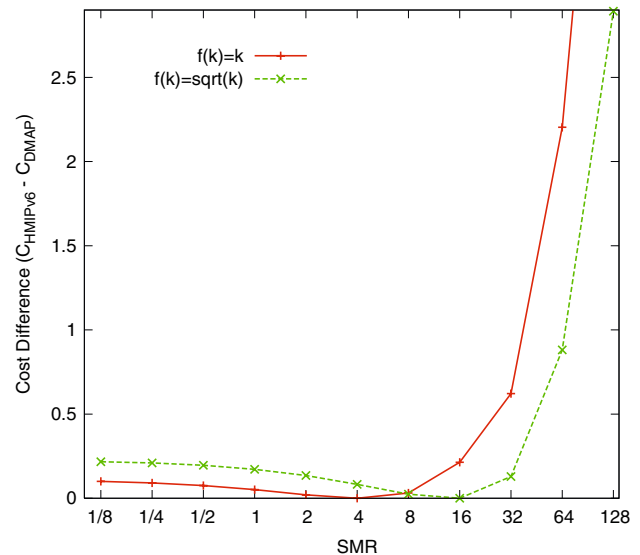**Fig. 4** Effect of α and β on $C_{\text{HMIPv6}} - C_{\text{DMAP}}$

[13]. We have tested the sensitivity to the form of $F(k)$. Figure 5 shows the effect of $F(k)$ on the cost difference $C_{\text{HMIPv6}} - C_{\text{DMAP}}$ for $F(k) = \sqrt{k}$ and $F(k) = k$. The cost difference curves exhibited are very similar in shape and are not sensitive to the form of $F(k)$. We conclude that all the conclusions drawn earlier from the case $F(k) = \sqrt{k}$ are valid.

6 Applicability and conclusion

In this paper, we proposed a novel DMAP scheme for integrated mobility and service management.

The core of the idea lies in allowing intelligent MNs to determine the best service areas for service handoffs to minimize the network signaling cost due to mobility and service management. We devised a procedure to compute the optimal service area size that would minimize the overall communication cost, when given a set of parameters characterizing the mobile node's mobility and service characteristics. We compared our scheme with basic MIPv6 and HMIPv6. Our scheme outperforms both basic MIPv6 and HMIPv6 in terms of the network communication overhead, the effect of which is especially pronounced when the service

Fig. 5 Effect of $F(k)$ on $C_{\mathrm{HMIPv6}} - C_{\mathrm{DMAP}}$

to mobility ration (SMR) is low for basic MIPv6, and either low or high for HMIPv6.

To apply the analysis results in the paper, one can execute the computational procedure at static time to determine optimal K_{opt} over a possible range of parameter values. Then a MN can perform a simple look-up operation to determine K_{opt} based on data collected at runtime. This allows each MN to dynamically determine the best service area size to minimize the overall communication cost. The performance gain is in the amount of communication cost saved per time unit per user, so the saving due to a proper selection of the best service area will have significant impacts since the cumulative effect for all mobile users over a long time period would be significant.

In the future, we plan to consider the implementation issue by building a testbed system to validate the analytical results as well as testing the sensitivity of the results with respect to other time distributions other than the exponential distribution used in the analysis. This testbed consists of DMAP-compliant MNs, ARs, and server applications in a Mobile IPv6 environment, leveraging existing code for HMIPv6 running on Linux platforms [8]. We will test scenarios with variations in the mobility and service characteristics of MNs in this testbed to compare DMAP with MIPv6 and HMIPv6 for experimental validation. We also plan to extend the research to the case in which a user proxy can

be used to execute the function of DMAP for database applications that use per-user, stateful proxies to effectively cache data items to support disconnected operations to further reduce the network communication cost.

References

1. Chen, I. R., & Verma, N. (2003). Simulation study of a class of autonomous host-centric mobility prediction algorithms for cellular and ad hoc networks. In *Proceedings of the 36th annual simulation symposium*, pp. 65–72. Orlando, USA.
2. Dimopoulou, L., Leoleis, G., & Venieris, I.S. (2004). Introducing a hybrid fast and hierarchical MIPv6 scheme in a UMTS-IP converged architecture. In *proceedings of the 29th IEEE international conference on local computer networks*, pp. 42–49. Tampa, USA.
3. Gu, B., & Chen, I.R. (2005). Performance analysis of location-aware mobile service proxies for reducing network cost in personal communication systems. *ACM Mobile Networks and Applications, 10*(4), 453–463.
4. Gustafsson, E., Jonsson, A., & Perkins, C. (2006). *Mobile IPv4 regional registration*. http://www.ietf.org/internet-drafts/draft-ietf-mip4-reg-tunnel-02.txt, IETF, Work in Progress.
5. Johnson, D., Perkins, C., & Arkko, J. (2004). *Mobility support in IPv6*. http://www.ietf.org/rfc/rfc3775.txt, IETF, Work in Progress.
6. Misra, A., Das, S., Dutta, A., McAuley, A., & Das, S.K. (2001). IDMP-based fast handoffs and paging in IP-based cellular networks. In *Proceedings of the International conference on 3G wireless and beyond*, pp. 427–432. San Francisco, USA.

7. Misra, A., Das, S., Dutta, A., McAuley, A., & Das, S.K. (2000). *IDMP: An Intra-Domain Mobility Management Protocol using Mobility Agents*. draft-mobileipmisra-idmp-00.txt, IETF, Work in Progress.
8. Nelson, R., Daley, G., & Moore, N. (2002). Implementation of hierarchical mobile IPv6 for linux. In *Proceedings of the 6th international symposium on communications interworking (IFIP Interworking 2002)*. Perth, Australia.
9. Pack, S., Choi, Y., & Nam, M. (2006). Design and analysis of optimal multi-level hierarchical Mobile IPv6 networks. *Wireless Personal Communications, 36*(2) 95–112.
10. Soliman, H., Castelluccia, C., El-Malki, K., & Bellier, L. (2005). Hierarchical mobile IPv6 mobility management. *http://www.ietf.org/rfc/rfc4140.txt, IETF, Work in Progress*.
11. Trivedi, K.S., Ciardo, G., & Muppala, J. (1999). *SPNP Version 6 User Manual*. Dept. of Electrical Engineering, Duke University, Durham, NC.
12. Xie, J. & Akyildiz, I.F. (2002). A novel distributed dynamic location management scheme for minimizing signaling costs in Mobile IP. *IEEE Transactions on Mobile Computing, 1*(3), 163–175.
13. Zhang, X., Castellanos, J., & Campbell, A. (2002). P–MIP: Paging extensions for Mobile IP. *ACM Mobile Networks and Applications, 7*(2), 127–141.

Weiping He received his BE degree from Huangzhong University of Science and Technology, China, in 1996, and his Master degree in Computer Science from Virginia Tech in 2002. He is a doctoral candidate in the Department of Computer Science at Virginia Tech. His research interests include future mobile communication system architectures, design and evaluation of mobility and service management schemes in mobile computing environments.

Baoshan Gu received the BS degree from University of Science and Technology of China, Hefei, China, in 1992 and the MS degree in computer science from Institute of Computing Technology, Chinese Academia of Science, Beijing, China, in 1995. He received his PhD degree in Computer Science from Virginia Tech in 2005. From 1995 to 2000, he was a research and development engineer in Institute of Computing Technology, Chinese Academia of Science. Since 2006, he has been a research scientist at NIH. His research interests include next-generation wireless system architectures, design and evaluation of location and service management schemes in mobile computing environments, and mobile multimedia systems.

Ing-Ray Chen received the BS degree from the National Taiwan University, Taipei, Taiwan, and the MS and PhD degrees in computer science from the University of Houston. He is a professor in the Department of Computer Science at Virginia Tech. His research interests include mobile computing, pervasive computing, multimedia, distributed systems, real-time intelligent systems, and reliability and performance analysis. Dr. Chen has served on the program committee of numerous conferences, including as program chair for 29th IEEE Annual International Computer Software and Application Conference in 2005, 14th IEEE International Conference on Tools with Artificial Intelligence in 2002, and 3rd IEEE Symposium on Application-Specific Systems and Software Engineering Technology in 2000. Dr. Chen currently serves as an editor for Wireless Personal Communications, The Computer Journal, and International Journal on Artificial Intelligence Tools. He is a member of the IEEE/CS and ACM.